1. Verify that $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

proof. We will show that $H(X,Y) = H(X) + H(Y|X)$ as follows:

$$H(X) + H(Y|X) = -\sum_x p_X(x)\log(p_X(x)) - \sum_{x,y} p_{X,Y}(x,y)\log\big(p_{Y|X}(y|x)\big) \quad (1)$$

Recall that $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$. Substituting and making use of the properties of the logarithm, the right-hand summation of (1) is

$$-\sum_{x,y} p_{X,Y}(x,y)\log(p_{X,Y}(x,y)) - p_{X,Y}(x,y)\log(p_X(x)) \quad (2)$$

distributing the double sum inwards and marginalizing gives us

$$\sum_{x,y} p_{X,Y}(x,y)\log(p_X(x)) = \sum_x p_X(x)\log(p_X(x)) \quad (3)$$

substituting and cleaning up terms gives us that

$$H(X) + H(Y|X) = H(X) + (H(X,Y) - H(X)) = H(X,Y), \quad (4)$$

as desired. $H(X,Y) = H(Y) + H(X|Y)$ follows via the same exact steps.

2. Extend the result of the previous exercise and to prove the following chain rule for entropy:

$$H(X_1,\ldots,X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1},\ldots,X_1) \quad (5)$$

proof. We will proceed via induction on n. Ignoring the trivial case where $n = 1$, we omit a proof of the base case when $n = 2$ as we can refer directly to the proof given above. Let us suppose that the following equality holds for all $k \leq n - 1$: $H(X_1, \ldots, X_{n-1}) = H(X_1) + H(X_2|X_1) + \ldots + H(X_{n-1}|X_{n-2}, \ldots, X_1)$. [a] Now, notice that

$$H(X_1, \ldots, X_n) = -\sum_{x_1 \ldots, x_n} p_{X_1,..,X_n}(x_1, .., x_n) \log(p_{X_1,..,X_n}(x_1, .., x_n)) \tag{6}$$

Now we refer to the [b] following:

$$p_{X_n|X_{n-1},\ldots X_1}(x_n|x_{n-1}, ..x_1) \cdot p_{X_1,..,X_{n-1}}(x_1, .., x_{n-1}) = p_{X_1,..,X_n}(x_1, .., x_n) \tag{7}$$

substituting into the logarithm in $H(X_1, \ldots, X_n)$, expanding, and then seperating into two terms we get the following two equalities:

$$-\sum_{x_1,..,x_n} p_{X_1,..,X_n}(x_1, .., x_n) \log\Big(p_{X_n|X_{n-1},\ldots X_1}(x_n|x_{n-1}, ..x_1)\Big) = H(X_n|X_{n-1}, \ldots, X_1) \tag{8}$$

and

$$-\sum_{x_1,\ldots,x_{n-1}} \log\big(p_{X_1,..,X_{n-1}}(x_1, .., x_{n-1})\big) \sum_{x_n} p_{X_1,..,X_n}(x_1, .., x_n) = H(X_{n-1}) \tag{9}$$

Therfore, we have split

$$H(X_1, \ldots, X_n) = H(X_n|X_{n-1}, \ldots, X_1) + H(X_{n-1}) \tag{10}$$

applying the inductive hypothesis on $H(X_{n-1})$ gives us that

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1}, \ldots, X_1), \tag{11}$$

as desired.

[a] The burden of proof is to now show $H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1}, \ldots, X_1)$.

[b] https://stats.stackexchange.com/questions/258379/why-is-pa-bc-pbc-pab-c

3. Prove that entropy is subadditive:

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i) \tag{12}$$

proof. We make use of the previous exercise and the fact that *conditioning does not increase entropy.*

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1}, \ldots, X_1) \tag{13}$$

Refer now to the following theorem:

Theorem 1. *The entropy $H(X)$ is greater than or equal to the conditional entropy $H(X|Y)$*

$$H(X) \geq H(X|Y) \tag{14}$$

Repeated application of the above theorem gives us $H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$, as desired.

4. Prove that entropy is additive when the random variables $X_1, X_2, \ldots, X_n$ are independent:

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) \tag{15}$$

proof.

$$H(X_1, \ldots, X_n) = -\sum_{x_1 \ldots, x_n} p_{X_1,..,X_n}(x_1, .., x_n) \log(p_{X_1,..,X_n}(x_1, .., x_n)) \tag{16}$$

Since the random variables are independent,

$$p_{X_1,..,X_n}(x_1, .., x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \ldots \cdot p_{X_n}(x_n) \tag{17}$$

Armed with this fact, we can substitute this expression into the logarithm, use the additive property of the logarithm to expand the summation, and finally, distribute the summations to marginalize the probabilities, giving us the equality $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i)$, as desired.
we will omit from explicitly showing the computations as they are essentially repeated applications of the techniques employed first two problems of this solution sheet.

Link to the template used